

**Kant's Grounding for the Metaphysics of Morals
A Very brief selective summary of sections I and II ***

**By
Geoffrey Sayre-McCord
UNC/Chapel Hill**

First Section

Kant begins the first section by distinguishing between things that are "good without qualification" or "unconditionally good" and things that are good, but only qualifiedly or under certain conditions. Although there are many things that fall into the second category -- everything that is good only because of its consequences (since its having those consequences is a condition on its being good) and even all but one thing that is good in itself. The one thing that is good without qualification, according to Kant, is (what he calls) a good will. Indeed, he claims, a good will is the only thing we can even imagine is good without qualification -- everything else being at best good only with qualification. Moreover, he maintains, the good will itself serves as a condition of the value of everything else -- something can be good only if it is (in some appropriate sense) compatible with a good will. In fact, "a good will seems" he claims, "to constitute the indispensable condition of being even worthy of happiness." (393-394)

If a good will is unconditionally good then its value, Kant points out, cannot depend upon its having good effects. For if its value did depend on its having good effects it would be valuable only on the condition that it had those effects. Take away the effects and you would take away the source of its value. Since its value is (by assumption only, so far) unconditional, it must then be valuable even absent its having any good effects. Its value must be contained within it. Kant supposes that we all have at least some idea of what he is referring to in speaking of a good will. Loosely speaking, it is the determination to do what, in effect, reason requires as right period. (394-394)

Kant recognizes that the idea that the role of reason is to make possible a good will rather than to help us satisfy our inclinations or make ourselves happy may seem highminded nonsense. So first he argues that if nature's purpose in giving us reason was to help us satisfy our inclinations or desires or preferences or to make us happy, it would have made a big mistake. Reason is ill suited to the task. In any case, the best way to show that the idea is not nonsense is to develop it in a way that might make it intelligible and Kant sets about doing just that. (395-396)

To understand what a good will is Kant turns our attention to the difference between a person (merely) doing her duty and her doing it because it is her duty. A good will finds its expression only in the latter case. Clearly, Kant points out,

a person is not exercising a good will when she does what she knows is wrong. But even when she is doing what she knows to be right she will be exercising a good will only if she does what she does because it is right and not because, say, she expects some reward or happens to want to do it. A person exercises a good will when, but only when, how that person acts is governed by whether so acting is compatible with her duty. (397-399)

The value of the action a person performs, insofar as it is an expression of good will, finds its value or worth "not in the purpose that is to be attained by it" (i.e., not in the consequences it might produce) but in the "maxim according to which the action is determined" (i.e., in the reason the agent had for acting in that way -- the recognition that so acting was at least compatible with duty). (399-400)

But if the value of an action done from duty is found not in the consequences it produces but in the respect for duty it expresses then one's duty must be to express that respect rather than to produce any particular effects. Doing one's duty because it is one's duty must then be a matter not of trying to achieve some effect but of conforming one's will to a principle of duty (law) that commands respect. (400-401)

What sort of principle might this be which can determine the will without appealing to some expected consequences of acting as it requires? Kant claims it must be a principle that requires one act so that the reasons one has for so acting could themselves be principles of the will. Moral action, it turns out, is in a very deep sense principled action -- action done on the basis of considerations that could themselves stand as principles. (401-402)

With this view in mind, Kant offers an example of a person considering whether or not to make a false promise. In asking oneself whether one may make such a promise one might be asking, Kant points out, whether so acting will promote one's welfare or one might be asking whether so acting is compatible with duty. The first question is often tricky to answer since it turns on so much that might be undiscoverable. The second question, though, is more readily answered Kant maintains because it is really the question of whether one could consistently commend the considerations that would lead one to make a lying promise as principles that might guide everyone's action. That we could not commend the considerations as principles of action, Kant holds, shows that acting on them is immoral, whatever the consequences might be. (402-405)

Second Section

The first section was given over to identifying and articulating our concept of duty, which demands and finds expression in a person's having a good will (i.e. in her doing her duty because it is her duty rather than because she expects some advantage or happens to feel like it). This is a concept with which we are all familiar, Kant maintains. Yet, while we all have experience of this concept, Kant argues that the concept itself is not one we get from experience. It is, in

* Copyright © 2000 by Geoffrey Sayre-McCord

his terminology, an *a priori*, not an *a posteriori*, concept. However we come to acquire the concept of duty, it is not by first experiencing instances of duty and then extracting our concept from those instances. Central to his argument is Kant's observation that morality's demands are unconditional, apply to all rational beings, and allow no exceptions. Whatever experiences we might have, he holds they couldn't be such as to validate by themselves any such demands. At the same time, he points out, in order to use examples as examples of someone acting morally we would already have to have the concept in order to determine whether the cases were appropriate -- i.e., actually cases of someone acting morally. Thus identifying the relevant cases presupposes our already possessing the concept that was purportedly derived from the cases. Moreover, he argues, to try to defend one's moral views by appeal to popular opinion or sentiment or an appeal to experience only will always result in clouded view of morality. (406-411)

If moral concepts "have their seat and origin completely a priori in reason," though, they must be understandable wholly without appeal to aspects of the human condition that can be known only from experience. No doubt moral concepts can be applied to people and the situations they face only in light of specific information experience provides, but the concepts themselves must be comprehensible independent of knowledge provided by experience. (412)

In order, then, to understand moral concepts conceived as applying to all rational beings regardless of their particular and contingent circumstances, the place to start (Kant holds) is with the concept of a rational being. The distinctive and defining features of rational beings, according to Kant, is their capacity to act as they do because of their "conception of laws" -- that is, their capacity not merely to be pushed around by forces but to act as they think they should (a capacity that involves being governed by their representation of some option as good or required). In a perfectly rational being, the representation of something as good or required is by itself, and without resistance, sufficient for action. For other beings -- such as human beings -- whose rational capacities govern a will that might be moved by various incentives, temptations, and fears, the representation of something as good or required is not, by itself, sufficient for action. In such beings, the determination by their reason that some option is good or required presents itself as a kind of command -- as the judgment that they ought to act accordingly (even if they want not to). (412-414)

Importantly, one's judgments that one ought to act in one way or another fall into two different categories. Sometimes the grounds one has for judging one ought so to act depend upon certain conditions being satisfied, so that the imperatives are hypothetical or conditional -- their practical force (i.e. their implications for action) depends upon the conditions in fact being satisfied. (The relevant conditions, it turns out, are that one has adopted a particular end or purpose.) Other times, the grounds one has for judging one ought so to act depend on nothing contingent, so that the imperatives are categorical -- their practical force (i.e. their implications for action) is unconditional and so not

dependent on the hypothesis that certain conditions are satisfied. As various hypothetical imperatives make clear, the judgments at issue here need not be moral in their ground or upshot. One might, for instance, judge that one ought not take another bite of the cake (despite the temptation) on the ground that one has decided to lose weight and foregoing such pleasures is a necessary step to achieving one's purpose. One needn't think one has a moral duty to lose weight, nor a moral duty not to eat given that one has decided to lose weight, even as it is clear that eating under such circumstances -- unless one abandons the diet -- would be irrational. Hypothetical imperatives in effect represent some action as good or required on that condition that one has some purpose, where the purpose is one that is, from the point of view of rationality, optional. Categorical imperatives, in contrast, in effect represent some action as good or required unconditionally. (414-415)

Hypothetical imperatives can be divided into those that are 'problematic' and those that are 'assertoric' depending upon whether the purpose in play is one that merely might be adopted or is one that has actually been adopted. Among the assertoric hypothetical imperatives, according to Kant, is the imperative that one act prudently (since doing so is required in order to achieve an end he believes we all necessarily have -- the end of promoting our own welfare). These hypothetical imperatives, whether problematic or assertoric, differ significantly from the imperatives of morality precisely in their conditionality, in their force depending on people having certain purposes. Morality requires that we act in certain ways not as a means to achieving some further end or purpose but directly and unconditionally, whether or not acting as it requires happens to further our ends or not. (415-417)

Against this background Kant turns to the question of how reason can *require* anything of us. What makes it true, in the case of hypothetical imperatives, that given certain ends or purposes it would be irrational not to act in a certain way? What might make it true, in the case of categorical imperatives, that it would be irrational not to act in a certain way regardless of one's ends? (417)

When it comes to hypothetical imperatives, Kant thinks there is no real problem in explaining the irrationality. As he sees it, to adopt an end is to set oneself to be the cause of achieving it, and the idea of one's being the cause of its achievement contains within it the idea of one's doing whatever is necessary. So one cannot rationally think of oneself as the cause and not be thinking of oneself as taking those necessary steps. To will the end (that is, to set oneself to achieving it) is thus to will the necessary means. (417) As Kant puts it: "it is one and the same thing to conceive of something as an effect that is possible in a certain way through me and to conceive of myself as acting in the same way with regard to the aforesaid effect." This might be a helpful parallel. To think of oneself as a bachelor is *ipso facto* to think of oneself as unmarried. The idea of one's being a bachelor contains within it the idea of one's being unmarried. So one cannot rationally think of oneself as a bachelor and not be thinking of

oneself as being unmarried. To believe one is to believe the other. In neither case, of course, is Kant committed to denying that people sometimes fail to will -- or believe -- as rationality requires. Such failures are all too familiar. What he is trying to do is identify what the failure consists in, and his answer is that (outside of morality) failures of practical rationality are failures to take what one recognizes to be the necessary steps to achieving one's ends while not abandoning the ends. In the face of a hypothetical imperative, there are two ways to preserve one's practical rationality: (1) take the necessary means to achieving the ends the give the imperative its practical force or (2) abandon those ends. Either is a rational response to one's situation. What is irrational is retaining the end while failing to take the steps one recognizes as necessary for its achievement. (417-420)

When it comes to categorical imperatives, Kant thinks the problem of making sense of even its possibility is a genuine challenge. Categorical imperatives differ from hypothetical imperatives in that, by not presupposing any particular ends they cannot derive their rational force from our having adopted those ends. Nor can we hope to discover a categorical imperative's force by looking to examples of agents exercising a good will because we can never be sure we have an example on hand. Instead of directly defending the force of the categorical imperative, Kant turns to the problem of identifying its content. What is it that reason might require unconditionally of all rational agents without regard to their particular ends or purposes? (420)

In answering this question, Kant notes that the content of a categorical imperative must be found solely in the form of the imperative (as an unconditional law that applies to all) and its demand that the will conform to it. "Hence," he writes as if the implication were at all clear, "there is only one categorical imperative and it is this: *Act only according to that maxim whereby you can at the same time will that it should become a universal law*" [p. 30; italics added]. More colloquially, the demand is that in whatever you do you should act for reasons that could serve as acceptable reasons for anyone. To make an exception of oneself, to act for reasons one could not acknowledge as acceptable were someone else to act on them, is to act immorally. How exactly this specific demand is supposed to follow from the form and force of a categorical imperative is a matter of serious and interesting controversy. Yet the principle itself has struck many as capturing an important feature of morality -- that its demands are universal and require, in some significant sense, that we treat all people the same. To act immorally is, on this view, to put oneself "above the law" -- not the civil law of some society or other (which might, after all, be a deeply immoral society) but the law one's own reason imposes upon one. And the immorality is found not principally in the acts one performs but, as Socrates emphasized, in the reasons one has for so acting. (421)

Kant then offers four examples of immoral actions, selected in part because they highlight the two different ways a person might fail to be acting on a maxim that would meet the requirements of the categorical imperative. A

maxim might fail by not being universalizable -- by being such that the very conception of it as being a law governing all is inconsistent. (We have, as Kant would put it, a perfect duty to refrain from acting on such maxims.) Or a maxim that might be universalizable might fail the requirements of the categorical imperative by being such that a person could not consistently will that the maxim be a universal law. (We have, as Kant would put it, an imperfect duty to refrain from acting on such maxims.) In both cases, the failure of the maxim is a failure of consistency in an important sense. There is no question that an immoral maxim can itself be perfectly consistent, after all people actually act on them. What is inconsistent is either (i) *the conception* of that maxim as a universal law or (ii) *willing* that the maxim serve as a universal law. Thus in testing a maxim (and so evaluating an action that might be performed on its basis) we can look for two kinds of inconsistency -- inconsistency in conception and inconsistency in willing.

The first and second examples are supposed to be cases in which the agent is considering acting on a maxim that cannot consistently be conceived as a universal law. The third and fourth examples are supposed to be cases in which the agent is considering acting on a maxim that can consistently be conceived as a universal law but that cannot consistently be willed. The difference between the first and third, on the one hand, and the second and the fourth on the other, is that the former have to do with one's treatment of oneself, while the latter have to do with one's treatment of others. (421-425)

Kant next offers some important observations concerning what could not serve as the source or foundation of morals. In particular, he argues, morality cannot legitimately be grounded merely in human nature, even as it applies to humans and applies in a way that is sensitive to our nature. (425-426)

The unconditional force of the categorical imperative is explicable, Kant maintains, only on the supposition that it derives its authority purely from our capacity, as rational beings, to determine our actions by the representation of certain options as required by rationally necessary laws (i.e. hypothetical and categorical imperatives). If those rationally necessary laws requires some action unconditionally, that action will be required of all rational beings (human and otherwise), and the actions prescribed are "objective ends" that are "valid for all rational beings". They will be the actions that are morally required. If, on the other hand, the rationally necessary laws require the adoption of some action only conditionally (say on the condition the agent happens to adopt some end or purpose) that action's worth is derived from, and contingent on, the agent having adopted the end or purpose in question. And assuming the action is not incompatible with a morally required action it will be morally permissible. (426-428)

How must we conceive of rational beings in order to make sense of their wills as grounds of unconditional (and conditional) value? Only, Kant supposes, by thinking of rational beings as ends in themselves (and not merely more or

less useful means to achieving some end a person might adopt). The value of everything other than rational beings, Kant goes on to say, is conditional and, in particular, conditional upon the rational beings happening to value those things. (428)

According to Kant, people inevitably view themselves (insofar as they are rational) as ends in themselves and as sources of value. Yet the reason each person has for viewing herself this way is equally a reason for her to value others similarly. So our viewing ourselves in this way commits us to viewing others too as ends in themselves. Thus reason requires that you "act in such a way that you treat humanity, whether in your own person or in the person of another, always at the same time as an end and never simply as a means" (p. 36). This is the second formulation of the categorical imperative. (429)

Using the same four examples Kant sets out to illustrate the two different ways in which this formulation of the categorical imperative rules out certain actions as immoral. The first two examples are of actions that involve a person treating either himself or another merely as a means. The second two examples, in contrast, don't involve a person treating someone as a mere means but do involve a failure actively to embrace others as ends, as beings with a worth beyond that conferred by their utility. (429-431)

It is worth noting that the second formulation of the categorical imperative does not rule out treating people as means -- using them to help further your own ends. What it rules out is treating them *merely* as means, as if their worth depends solely on their serviceability.

The over-all picture is that anything whatsoever a rational being might decide to do is morally permissible *as long as* in deciding to do it the agent is respecting the categorical imperative -- that is, acting for reasons the agent could will to be universal laws and (Kant thinks equivalently) treating all rational beings involved as ends and not merely means.